# APPLICATION

# FOR

# UNITED STATES LETTERS PATENT

**TITLE:** **AUDIO SIGNAL PROCESSING**

**APPLICANT:** **J. RICHARD AYLWARD**

"EXPRESS MAIL" Mailing Label Number EM040829440US

Date of Deposit 12-24-98

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as **"Express Mail Post Office To Addressee"** with sufficient postage on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

Douglas W. Roache

Douglas W. Roache

AUDIO SIGNAL PROCESSING

The invention relates to processing audio signals, and more particularly to processing one or more audio input signals to provide more audio signals.

It is an important object of the invention to provide an audio signal processing system to provide a plurality of audio channel output signals from one or more input signals.

According to the invention, a method for processing a single channel audio signal to provide a plurality of audio channel signals includes separating the single channel audio signal into a first separated signal characterized by a frequency spectrum generally characteristic of speech, and a second separated signal; generating a first channel signal from the first separated signal; and modifying the second separated signal to produce the remainder of the plurality of channel signals.

In another aspect of the invention, an audio signal processing apparatus for processing a single channel audio signal to provide a plurality of audio channel signals, includes a speech separator for separating the audio signal into a first separated signal characterized by a frequency spectrum generally characteristic of speech, and a second separated signal; and a circuit coupled to the speech separator for generating a first subset of the plurality of audio channel signals from the second separated signal,.

In another aspect of the invention, an audio signal processing system includes an input terminal for a single input channel signal; a center channel output terminal for a center channel output signal *C*; a plurality of output terminals for a corresponding plurality of output channel signals; a speech separator inter-coupling the input terminal and the center channel output terminal for separating the single channel input signal into a speech audio signal and a nonspeech audio signal; and a circuit coupling the speech separator to the plurality of output terminals for providing, responsive to the nonspeech audio signal, a corresponding plurality of audio channel signals on the output terminals.

In another aspect of the invention, a method for processing a single channel audio signal to provide two decodable audio channel signals decodable into five audio channel signals includes separating the single channel audio signal into a first separated signal characterized by a frequency spectrum generally characteristic of speech, and a second separated signal; processing the first separated signal to provide a center channel signal $C$; modifying the second separated signal to provide a left channel signal $L$, a right channel signal $R$, a left surround channel signal $L_S$, and a right surround channel signal $R_S$; combining the center channel signal, a sum of the left surround and the right surround channel signals and the left channel signal to produce a first of the two decodable audio channel signals; and combining the center channel signal, a sum of the left surround and the right surround channel signals and the right channel signal to produce a second of the two decodable audio channel signals.

In another aspect of the invention, a method for processing a single channel audio signal to provide three decodable audio channel signals subsequently decodable into five audio channel signals, comprises separating the single channel audio signal into a first separated signal characterized by a frequency spectrum generally characteristic of speech, and a second separated signal; processing the first separated signal to provide a center channel signal, the center channel signal comprising the first decodable audio signal; modifying the second separated signal to provide a left channel signal, a right channel signal, a left surround channel signal, and a right surround channel signal; combining a sum of the left surround and the right surround channel signals and the left channel signal to produce a second of the three decodable audio channel signals; and combining a sum of the left surround and the right surround channel signals and the right channel signal to produce a third of the three decodable audio channel signals.

In another aspect of the invention, a method for processing two input audio channel signals to provide more than two output audio channel signals includes separating each of the two input audio channel signals into a first separated signal characterized by a frequency spectrum generally characteristic of speech, and a second separated signal; combining the first separated signal of the first input audio channel signal and the first separated signal of the second input audio channel signal to form a first of the more than

two output audio channel signals; transmitting the second separated signal of the first input signal as a second of the more than two output audio channel signals; and transmitting the second separated signal of the second input signal as a third of the more than two output channel signals.

In still another aspect of the invention, an audio signal processing apparatus for processing two input audio channel signals to provide more than two output audio channel signals includes a first speech separator for separating a first of the two input audio channel signals into a first separated signal characterized by a frequency spectrum characteristic of speech to provide a first of the more than two output audio channel signals; a second speech separator for separating a second of the two audio channel signals into a first separated signal characterized by a frequency spectrum characteristic of speech, and a second of the more than two output audio channel signals; and a combiner for combining the first and second separated signals to form a third of the more than two output audio channel signals.

Other features, objects, and advantages will become apparent from the following detailed description, which refers to the following drawings in which:

FIG. 1 is a block diagram of a single channel audio signal processing system according to the invention;

FIGS. 2a and 2b are circuit diagrams of circuits implementing the speech separator and the multichannel emulator of FIG. 1;

sub a¹

~~FIGS. 3a – 3c are block diagrams of alternate embodiments of the postemulation processing system of FIG. 1; and~~

FIG. 4 is a circuit diagram of a circuit implementing the principles of the invention in a two input channel system.

With reference now to the drawings and more particularly to FIG. 1, there is shown a single channel audio signal processing system according to the invention. Single channel signal input terminal 10 is connected to speech separator 12. Speech separator 12 is coupled to multichannel emulator 16 by nonspeech signal line 14 and is coupled to

postemulation processing system 20 by speech signal line 18. Multichannel emulator 16 is coupled to postemulation processing system 20 through emulated signal lines $22_a - 22_z$. Speech separator 12 has two output taps, speech level tap 26 and nonspeech level tap 28.

In operation, a single channel signal, such as a monophonic audio signal is input at input terminal 10. The single channel input signal is separated into a speech signal and a nonspeech signal by speech separator 12. The speech signal is output on line 18 as a first output channel signal to postemulation processing system 20. The nonspeech signal portion on line 14 is then processed by multichannel emulator 16 to produce multiple output audio channel signals, which are then processed by postemulation processing system 20. The elements and function of postemulation processing system 20 will be shown in more detail in FIGS. 3a – 3d and explained in more detail in the corresponding portion of the disclosure.

Speech separator 12 may include a bandpass filter in which the pass band is a frequency range, such as 300 Hz to 3 kHz, or such as the so-called "A Weighted" filter described in publication ANSI S1.4-1983, published by the American Institute for Physics for the Acoustical Society of America, which contains the range of frequencies or spectral components commonly associated with speech. Other filters having different characteristics may be used to account for different languages, intonations, and the like. Speech separator 12 may also include more complex filtering networks or some other sort of speech recognition device, such as a microprocessor adapted for recognizing signal patterns representative of speech.

An audio signal processing system according to FIG. 1 is advantageous because transmissions or sources (such as videocassettes) having monophonic audio tracks can be presented on five channel audio systems with realistic "surround" effect, including on-screen localization of dialog.

Referring now to FIG. 2a, there is shown one embodiment of a circuit implementing speech separator 12 and multichannel emulator 16. The circuit has a single input channel and five output channels. The input channel may be a monophonic audio

signal input, and the five output channels may be a left channel, a right channel, a left surround channel, a right surround channel and a center channel, as in a home theater system.

Speech separator 12 may include input terminal 10, which is coupled to the input terminal of speech filter 80, to a + input terminal of first signal summer 82 and to a + input terminal of second signal summer 84. The output terminal of speech filter 80 is coupled to first multiplier 55 and to speech level tap 26 and is coupled to the – input terminal of first signal summer 82. The output of first multiplier 55 is coupled to center channel signal line 22C and to the – input terminal of second signal summer 84. The output terminal of second signal summer 84 is coupled to multichannel emulator 16 through nonspeech content signal line 14. The output terminal of first signal summer 82 is coupled to nonspeech level tap 28.

Nonspeech content signal line 14 is coupled through delay unit 32 to a + input terminal of third signal summer 34, and a – terminal of fourth signal summer 36, thereby providing multiple paths for processing the nonspeech signal. The output terminal of delay unit 32 is coupled to a – input terminal of fourth signal summer 36, to a + input terminal of seventh signal summer 46 and a + input terminal of eighth signal summer 48. The output terminal of third signal summer 34 is coupled to an input terminal of fifth signal summer 38 and to an input terminal of second multiplier 40. The output terminal of fourth signal summer 36 is coupled to a + input terminal of sixth signal summer 42 and to an input terminal of third multiplier 44. The output terminal of fifth signal summer 38 is coupled to left channel signal line 22L and to a – input terminal of seventh signal summer 46. The output terminal of sixth signal summer 42 is coupled to right channel signal line 22R and to a + input terminal of eighth signal summer 48. The output terminal of seventh signal summer 46 is coupled to right surround channel signal line $22R_S$. The output terminal of eighth signal summer 48 is coupled to left surround signal line $22L_S$. The output terminal of delay unit 32 is coupled to an input terminal of seventh signal summer 46 and to an input terminal of eighth signal summer 48.

Delay unit 32 may apply a 5ms delay to the signal. Third signal summer 34 may scale input from delay unit 32 by a factor of 0.5. Fourth signal summer 36 may scale input from delay unit 32 by a factor of 0.5. Seventh signal summer 46 and eighth signal summer 48 may scale their outputs by a factor of 0.5. First multiplier 55 may multiply the input signal from speech filter 80 by a factor of $\frac{|C|}{|C|+|\overline{C}|}$ (hereinafter α) where $|C|$ is the time averaged magnitude of the speech signal on line 18 and $|\overline{C}|$ is the time averaged magnitude of the complement of the speech signal. $|C|$ and $|\overline{C}|$ may be measured at speech tap 26 and nonspeech tap 28, respectively. Time averaging of $|C|$ and $|\overline{C}|$ may be done over a sample period, such as 300ms. Time averaging of the value of $|C|$ may also be done over two different time periods, such as 300mS and 30mS, combined, and scaled. Multipliers 40, 44, may multiply their inputs by a factor of α.

For a monophonic input signal *M*, the circuit of FIG. 2a yields the following output signals at the following signal lines:

| Signal Line | Channel | Signal | Value as $\alpha \to 0$ | Value as $\alpha \to 1$ |
|---|---|---|---|---|
| 22C | Center | $\alpha C$ | 0 | C |
| 22L | Left (L) | $\overline{C}+.5\overline{C}\Delta t-\alpha(\overline{C}-.5\overline{C}\Delta t)$ | $\overline{C}+.5\overline{C}\Delta t$ | $\overline{C}\Delta t$ |
| 22R | Right (R) | $\overline{C}-.5\overline{C}\Delta t-\alpha(\overline{C}+.5\overline{C}\Delta t)$ | $\overline{C}-.5\overline{C}\Delta t$ | $-\overline{C}\Delta t$ |
| $22L_s$ | Left Surround | $.5(\overline{C}\Delta t+R)$ | $.5(\overline{C}+1.5\overline{C}\Delta t)$ | 0 |
| $22R_s$ | Right Surround | $.5(\overline{C}\Delta t-L)$ | $.5(-\overline{C}+1.5\overline{C}\Delta t)$ | 0 |

Table 1.

where $C$ represents the speech content of signal $M$, $\overline{C}$ represents the nonspeech content of signal $M$, $\overline{C}\Delta t$ represents the nonspeech content of signal $M$ delayed in time, $L$ represents the left channel signal, $R$ represents the right channel signal, and $\alpha$ is as defined above.

Referring now to FIG. 2b, there is shown a second embodiment of a circuit implementing speech separator 12 and multichannel emulator 16. The circuit includes single input channel and five output channels. The input channel may be a monophonic audio input, and the five output channels may be a left channel, a right channel, a left surround channel, a right surround channel and a center channel, as in a home theater system.

The circuit of FIG. 2b is substantially identical to the circuit of FIG. 2a, except that in FIG. 2b, the input of multiplier 55 is directly coupled to input terminal 10 rather than to the output of speech filter 80, and the signal on center channel signal line 22C is scaled by a factor of 1.414.

A circuit according to the invention is advantageous because it can provide realistic five channel effect from monophonic signals. In the left and right channels, the $\overline{C}$ components are in phase, but the $.5\overline{C}\Delta t$ components are out of phase, which results in a stereo effect. In the left surround and right surround channels, the $\overline{C}$ component are out of phase, which prevents localization on the left surround and right surround channels. The speech content of signal $M$ is radiated by the center channel only, and is scaled to provide the appropriate power level so that speech is localized on the screen and is of the appropriate level.

A circuit according to the invention is also advantageous because total signal power is maintained. As can be seen in the circuit if FIGS. 2a and 2b, and table 1, the variable gain a is directly applied to the signal in channel 22C and the signal $\alpha\left(\overline{C}+.5\overline{C}\Delta t\right)$ is subtractively combined with the signal in channels 22L and 22R so that increase in

variable gain a results in an increase in signal strength of the signal in channel 22C and a decrease in signal strength in the signals in channels 22L and 22R.

A circuit according to the invention is also advantageous of because the relative proportion of the sound radiated by speakers connected to the various channels is appropriate relative to the speech content of the monophonic input signal. If input signal $M$ contains no speech, then $C$ approaches zero, $\overline{C}$ approaches $M$, and $\alpha$ approaches zero. In this situation, there is no signal on the center channel and the signals on the other channels are as shown in Table 1. If signal $M$ is predominantly speech, then $C$ approaches $M$, $\overline{C}$ approaches zero, and $\alpha$ approaches one. In this case, the signal in the left and right surround channels approaches zero, and the signal on the left and right channels approaches $\overline{C}\Delta t$ and $-\overline{C}\Delta t$, respectively. Since the signal is delayed, the center channel is the source of first arrival information, and information from the complementary channels arrives later in time, so that a listener will localize on the radiation from the center channel. When the signal is predominantly speech, the signals on the left surround and right surround channels approach zero, so that there is no radiation from the surround speakers.

A further advantage of the circuit according to the invention is that the combining effect of the circuit is time-varying so that the perceived sources of the left and right channels are not spatially fixed.

Referring to FIGS. 3a – 3d, there are shown alternate embodiments of postemulation processing system 20. In FIG. 3a, signal lines 22L, $22L_S$, 22R, $22R_S$ and 22C may be coupled to respective electroacoustical transducers 52L, $52L_S$, 52R, $52R_S$, and 52C which radiate sound waves corresponding to the signals on signal lines 22L, $22L_S$, 22R, $22R_S$ and 22C, respectively. Electroacoustical transducers 52L, $52L_S$, 52R, $52R_S$, and 52C may be the left, left surround, right, right surround, and center channel speakers of a home theater system.

In the embodiment of FIG. 3b, postemulation processing system 20 may include a crossover network 54, which couples signal lines 22L, $22L_S$, 22R and $22R_S$ to tweeters respective tweeters 56L, $56L_S$, 56R, and $56R_S$ and to subwoofer 58 and signal line 22C

may be coupled to electroacoustical transducer 60. Tweeters 56L, $56L_S$, 56R, and $56R_S$ may be the left, left surround, right, and right surround speakers, subwoofer 58 may be the subwoofer, and electroacoustical transducer 60 may be the center channel of a subwoofer/satellite type home theater system.

In the embodiment of FIG. 3c, postemulation processing system 20 may include a circuit for downmixing the outputs of multichannel emulator 16 into three channel signals suitable for recording, transmission or for playback on a three-channel system. Input terminals of ninth signal summer 62 are coupled to signal lines $22L_S$ and $22R_S$. The output terminal of ninth signal summer 62 is coupled to an input terminal of tenth signal summer 64 and an input terminal of eleventh signal summer 66. Signal from ninth signal summer 62 to tenth signal summer 64 may be scaled by a factor of 0.707, and signal from ninth signal summer 62 to eleventh signal summer 66 may be scaled by a factor of -0.707. An input terminal of tenth signal summer 64 may be coupled to signal line 22L so that the output signal of tenth signal summer 64 is $0.707(L_S + R_S)+L$, (where $L_S$, $R_S$, and L represent the inputs from signal lines $22L_S$, $22R_S$, and 22L respectively) which is output at left channel output terminal 86L. Input of eleventh signal summer 66 may be coupled to signal line 22R so that the output of eleventh signal summer 66 is $-0.707(L_S + R_S)+R$, (where $L_S$, $R_S$, and R represent the inputs from signal lines $22L_S$, $22R_S$, and 22R respectively) which is output at right channel output terminal 86R. Signal line 22C is coupled to center channel output terminal 86C.

has been provided

In the embodiment of FIG. 3d, postemulation processing system 20 includes a circuit for downmixing the output signals of multichannel emulator 16 into two channel signals suitable for recording, transmission, or for playback on a two-channel system. Input terminals of signal summer 62 are coupled to signal lines $22L_S$ and $22R_S$. The output terminal of ninth signal summer 62 is coupled to an input terminal of tenth signal summer 64 and an input terminal of eleventh signal summer 66. Signal from ninth signal summer 62 to tenth signal summer 64 may be scaled by a factor of 0.707, and signal from ninth signal summer 62 to eleventh signal summer 66 may be scaled by a factor of -0.707. An input terminal of tenth signal summer 64 is coupled to signal line 22L so that the

output signal of tenth signal summer 64 is $0.707(L_S + R_S)+L$, (where $L_S$, $R_S$, and L represent the signals on signal lines $22L_S$, $22R_S$, and 22L respectively). The output terminal of tenth signal summer 64 is coupled to an input terminal of twelfth signal summer 68. An input terminal of eleventh signal summer 66 may be coupled to signal line 22R so that the output signal of eleventh signal summer 66 is $-0.707(L_S + R_S)+R$, (where $L_S$, $R_S$, and R represent the inputs from signal lines $22L_S$, $22R_S$, and 22R respectively). The output terminal of eleventh signal summer 66 is coupled to an input terminal of thirteenth signal summer 70. Signal from first multiplier 55 to tenth signal summer 68 may be scaled by a factor of 0.707, so that output signal of tenth signal summer 68 is $.707C+707(L_S + R_S)+L$, (where $L_S$, $R_S$, L, and C represent the inputs from signal lines $22L_S$, $22R_S$, and 22L and from first multiplier 55 respectively). The output terminal of tenth signal summer is coupled to left channel terminal output 84L. Signal from first multiplier 55 to thirteenth signal summer 70 may be scaled by a factor of 0.707, so that output of thirteenth signal summer 70 is $.707C-707(L_S + R_S)+L$, (where $L_S$, $R_S$, L, and C represent the inputs from signal lines $22L_S$, $22R_S$, 22L, and 22C, respectively). The output terminal of thirteenth signal summer 70 is coupled to right channel output terminal 84R.

The embodiments of FIGS. 3c and 3d are advantageous because they can be rerecorded or retransmitted in two- or three-channel format and subsequently decoded for presentation in five-channel format.

Referring now to FIG. 4, there is shown a circuit implementing the principles of the invention in a two input channel system. Left input channel terminal 90L is coupled to an input of left speech filter 92L and additively coupled with left summer 94L. The output of speech filter 92L is differentially coupled with an input of left summer 94L and additively coupled with center summer 96C. The output of left summer 94L is coupled with left channel output terminal 98L and left surround summer $94L_S$ and differentially coupled with right surround summer $94R_S$. Right input channel terminal 90R is coupled to an input of right speech filter 92L and additively coupled with right summer 94R. The output of speech filter 92R is differentially coupled with an input of right summer 94R and additively coupled with center summer 96C. The output of right summer 94R is coupled

with right channel output terminal 98R and right surround summer $94R_S$ and differentially coupled with left surround summer $94L_S$. The output of left surround summer $94L_S$ is coupled to left surround output terminal $98L_S$ and output of right surround summer $94R_S$ is coupled to right surround output terminal $98R_S$.

In operation a two-channel input signal, such as a stereophonic signal having left and right channels is input at input terminals 90L and 90R, respectively. The circuit separates the speech band portion of the signal, combines the left speech band portion $C_L$ and the right speech band portion $C_R$, combines them, and scales them to form a center channel signal which is output at center channel terminal 98C. The nonspeech portion of the left channel signal and the nonspeech portion of the right channel signal are output at left channel output terminal 98L and right channel output terminal 98R, respectively. The output of center channel terminal 98C may then be used as the center channel of a three- or five-channel audio system. The output of left channel output terminal 98L and right channel output terminal 98R can then be used as the left and right channels of a three channel system. If a five channel output is desired, the output of summer 94R may be differentially combined with the output of summer 94L and scaled to form the left surround channel signal which is output at left surround output terminal $98L_S$,and the output of summer 94L may be differentially combined with the output of summer 94R and scaled to form the right surround channel signal which can be output at the right surround output terminal $98R_S$.

Other embodiments are within the claims.